BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 6: e94196
doi: 10.3897/biss.6.94196

OPEN ACCESS

Conference Abstract

Establishing Plankton Imagery Dataflows Towards International Biodiversity Data Aggregators

Patricia Martin-Cabrera[‡], Ruben Perez Perez[‡], Jean-Olivier Irrison[§], Fabien Lombard[§], Klas Ove Möller[|], Saskia Rühl[|], Veronique Creach[¶], Markus Lindh[#], Lars Stemmann[§], Lennert Schepers[‡]

‡ Flanders Marine Institute, Ostend, Belgium
§ Institut de la Mer de Villefranche, Centre National de la Recherche Scientifique – Sorbonne Université, Villefranche sur-Mer, France
| Institute of Carbon Cycles, Helmholtz-Zentrum Geesthacht, Hereon, Geesthacht, Germany
¶ Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, United Kingdom
# Swedish Meteorological and Hydrological Institute, Västra Frölunda, Sweden

Corresponding author: Patricia Martin-Cabrera (patricia.cabrera@vliz.be)

Received: 29 Aug 2022 | Published: 07 Sep 2022

Citation: Martin-Cabrera P, Perez Perez R, Irrison J-O, Lombard F, Ove Möller K, Rühl S, Creach V, Lindh M, Stemmann L, Schepers L (2022) Establishing Plankton Imagery Dataflows Towards International Biodiversity Data Aggregators. Biodiversity Information Science and Standards 6: e94196.
https://doi.org/10.3897/biss.6.94196

Abstract

Over the last decade, plankton research has experienced extensive developments in automatic image acquisition for identifying and quantifying plankton species. This information is useful for the reporting of plankton occurrences and ecological data. Imagery instruments can vary in the way they sample (benchtop or *in situ* imagers) and the particle's size range they target (see Lombard et al. (2019) for an extensive comparison of instruments and specifications). However, due to the wide variety of instruments and their (automatic) output data and formats, it is challenging to integrate datasets that originate from different sources. For this reason, we developed recommendations for plankton imagery data management, which can promote the ability to make these datasets as FAIR (Findable, Accessible, Interoperable and Reusable principles), as possible. The workflow presented here could inspire other Biodiversity Information Standards TDWG communities working with (automated) imagery data (e.g., camera traps) such as the Audubon Core and Machine Observations Interest Group.

© Martin-Cabrera P et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The recommended data format follows the OBIS-ENV-DATA format (De Pooter et al. 2017), a Darwin Core-based approach to standardise biodiversity data (Wieczorek et al. 2012) used in EurOBIS, the European node of the Ocean Biodiversity Information System (OBIS) and EMODnet Biology, the European Marine Biodiversity Data Network. However, this format does not include sufficient information for imagery data, therefore we propose the use of additional Darwin Core terms. For example, by including the terms identifiedBy, identificationVerificationStatus and identificationReferences in the Occurrence table, more clarity is reported regarding the uncertainty of the classification made by an algorithm. Thus, data providers can publish manually validated datasets or datasets produced by fully automated plankton identification workflows; and users can choose to use validated or not validated data. See in Suppl. material 1 a practical example on how to report an imagery dataset following the best practices.

Moreover, the OBIS-ENV-DATA format allows the ingestion of additional information thanks to the use of the Darwin Core (DwC) Extended Measurement Or Facts or eMoF extension in the DwC Event core. The eMoF stores biotic, abiotic and sampling measurements and facts that are related to the Event and Occurrence table. An important aspect of this extension is that it includes standardised terms and controlled vocabularies, such as the British Oceanographic Data Centre (BODC) vocabularies, to standardise parameters that are not covered by DwC. The advantages of these is to unambiguously report information and to include those measurements that cannot be reported in the Event and Occurrence table (e.g., reporting abundance or biomass of plankton), and that are crucial to investigate ecosystem functioning questions. As a consequence, biodiversity data aggregators can extend their scope beyond species occurrence data.

Fig. 1 summarises a typical dataflow that goes from imagery data acquisition to publication in several steps:

1. Images are cropped and classified with software. This can be done in EcoTaxa, a web application that allows users to taxonomically classify images of individual organisms.

2. Data is formatted in OBIS-ENV-DATA format. This format can be exported from EcoTaxa through its API.

3. Data is submitted to EurOBIS via the IPT (Integrated Publishing Toolkit).

4. Data is quality controlled by the BioCheck tool.

5. Data in EurOBIS can flow to EMODnet Biology, OBIS and GBIF (Global Biodiversity Information Facility).

Plankton imagery instrument operators now have the possibility to format their data following the best practices and recommendations for plankton imagery data management (Martin-Cabrera et al. 2022). After a dataset is formatted following these guidelines, it can be submitted to the international biodiversity data aggregators, EurOBIS, EMODnet Biology and GBIF. Additionally a (semi) automated dataflow is presented where data

providers can classify images in EcoTaxa and export the data in the required formats using an API before submission to EurOBIS.

The next steps are to disseminate these best practices, encouraging plankton imagery data generators to implement these workflows to share their data easily, enriching these data portals and encouraging cross collaborations to create data products covering broader geographic scales and plankton species.

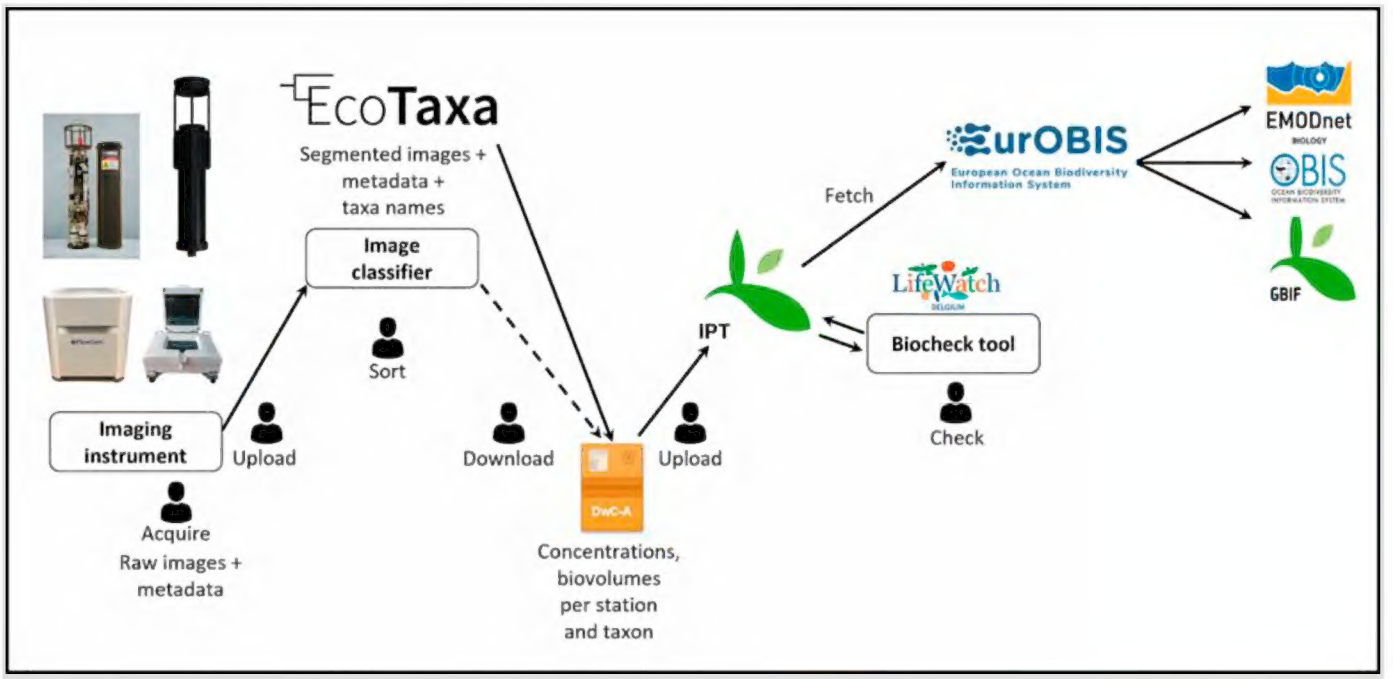

Figure 1.

Complete workflow from imagery instrument to international biodiversity data aggregators.

Keywords

imagery instruments, imagery data, Darwin Core, best practices

Presenting author

Patricia Martin-Cabrera

Presented at

TDWG 2022

Acknowledgements

This work was supported by the JERICO-S3 project, that received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement no. 871153.

Hosting institution

Flanders Marine Institute

References

- De Pooter D, Appeltans W, Bailly N, Bristol S, Deneudt K, Eliezer M, Fujioka E, Giorgetti A, Goldstein P, Lewis M, Lipizer M, Mackay K, Marin M, Moncoiffé G, Nikolopoulou S, Provoost P, Rauch S, Roubicek A, Torres C, van de Putte A, Vandepitte L, Vanhoorne B, Vinci M, Wambiji N, Watts D, Klein Salas E, Hernandez F (2017) Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5: e10989. https://doi.org/10.3897/bdj.5.e10989
- Lombard F, Boss E, Waite A, Vogt M, Uitz J, Stemmann L, Sosik H, Schulz J, Romagnan J, Picheral M, Pearlman J, Ohman M, Niehoff B, Möller K, Miloslavich P, Lara-Lpez A, Kudela R, Lopes R, Kiko R, Karp-Boss L, Jaffe J, Iversen M, Irisson J, Fennel K, Hauss H, Guidi L, Gorsky G, Giering SC, Gaube P, Gallager S, Dubelaar G, Cowen R, Carlotti F, Briseño-Avena C, Berline L, Benoit-Bird K, Bax N, Batten S, Ayata SD, Artigas LF, Appeltans W (2019) Globally Consistent Quantitative Observations of Planktonic Ecosystems. Frontiers in Marine Science 6: 196 https://doi.org/10.3389/fmars.2019.00196
- Martin-Cabrera P, Perez Perez R, Irisson J, Lombard F, F. MO, Rühl S, Creach V, Lindh M, Stemmann L, Schepers L (2022) Best practices and recommendations for plankton imagery data management: Ensuring effective data flow towards European data infrastructures. Flanders Marine Institute, VLIZ, Ostend. 31 pp. https://doi.org/10.25607/OBP-1742
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1):29715. https://doi.org/10.1371/journal.pone.0029715

Supplementary material

Suppl. material 1: Imagery dataset example doi

Authors: Patricia Martin-Cabrera
Data type: Imagery data
Brief description: Practical example of the data format proposed for plankton imagery data
Download file (251.34 kb)